

# WALD TESTS OF SINGULAR HYPOTHESES

BY MATHIAS DRTON AND HAN XIAO

*University of Washington and Rutgers University*

Motivated by the problem of testing tetrad constraints in factor analysis, we study the large-sample distribution of Wald statistics at parameter points at which the gradient of the tested constraint vanishes. When based on an asymptotically normal estimator, the Wald statistic converges to a rational function of a normal random vector. The rational function is determined by a homogeneous polynomial and a covariance matrix. For quadratic forms and bivariate monomials of arbitrary degree, we show unexpected relationships to chi-square distributions that explain conservative behavior of certain Wald tests. For general monomials, we offer a conjecture according to which the reciprocal of a certain quadratic form in the reciprocals of dependent normal random variables is chi-square distributed.

**1. Introduction.** Let  $f \in \mathbb{R}[x_1, \dots, x_k]$  be a homogeneous  $k$ -variate polynomial with gradient  $\nabla f$ , and let  $\Sigma$  be a  $k \times k$  positive semidefinite matrix with positive diagonal entries. In this paper, we study the distribution of the random variable

$$(1.1) \quad W_{f,\Sigma} = \frac{f(X)^2}{(\nabla f(X))^T \Sigma \nabla f(X)},$$

where  $X \sim \mathcal{N}_k(0, \Sigma)$  is a normal random vector with zero mean and covariance matrix  $\Sigma$ . The random variable  $W_{f,\Sigma}$  arises in the description of the large-sample behavior of Wald tests with  $\Sigma$  being the asymptotic covariance matrix of an estimator and the polynomial  $f$  appearing in a Taylor approximation to the function that defines the constraint to be tested.

In regular settings, the Wald statistic for a single constraint converges to  $\chi_1^2$ , the chi-square distribution with one degree of freedom. This familiar fact is recovered when  $f(x) = a^T x$ ,  $a \neq 0$ , is a linear form and

$$(1.2) \quad W_{f,\Sigma} = \left( \frac{a^T X}{\sqrt{a^T \Sigma a}} \right)^2$$

becomes the square of a standard normal random variable; the vector  $a$  corresponds to a nonzero gradient of the tested constraint. Our attention is

---

*AMS 2000 subject classifications:* 62F05, 62E20

*Keywords and phrases:* Asymptotic distribution, factor analysis, large-sample theory, singular parameter point, tetrad, Wald statistic

devoted to cases in which  $f$  has degree two or larger. These singular cases occur when the gradient of the constraint is zero at the true parameter.

For likelihood ratio tests, a large body of literature starting with Chernoff (1954) describes large-sample behavior in irregular settings; examples of recent work are Azaïs, Gassiat and Mercadier (2006), Drton (2009), Kato and Kuriki (2013) and Ritz and Skovgaard (2005). In contrast, much less work appears to exist for Wald tests. Three examples we are aware of are Glonek (1993), Gaffke, Steyer and von Davier (1999) and Gaffke, Heiligers and Offinger (2002) who treat singular hypotheses that correspond to collapsibility of contingency tables and confounding in regression. Our own interest is motivated by the fact that graphical models with hidden variables are singular (Drton, Sturmfels and Sullivant, 2009, Chap. 4).

In graphical modeling, or more specifically in factor analysis, the testing of so-called ‘tetrad constraints’ is a problem of particular practical relevance (Bollen, Lennox and Dahly, 2009; Bollen and Ting, 2000; Hipp and Bollen, 2003; Silva et al., 2006; Spirtes, Glymour and Scheines, 2000). This problem goes back to Spearman (1904); for some of the history see Harman (1976). The desire to better understand the Wald statistic for a tetrad was the initial statistical motivation for this work. We solve the tetrad problem in Section 5; the relevant polynomial is quadratic, namely,  $f(x) = x_1x_2 - x_3x_4$ . However, many other hypotheses are of interest in graphical modeling and beyond (Drton, Sturmfels and Sullivant, 2007; Drton, Massam and Olkin, 2008; Sullivant, Talaska and Draisma, 2010; Zwiernik and Smith, 2012). In principle, any homogeneous polynomial  $f$  could arise in the description of a large-sample limit and, thus, general distribution theory for the random variable  $W_{f,\Sigma}$  from (1.1) would be desirable.

At first sight, it may seem as if not much concrete can be said about  $W_{f,\Sigma}$  when  $f$  has degree two or larger. However, the distribution of  $W_{f,\Sigma}$  can in surprising ways be independent of the covariance matrix  $\Sigma$  even if  $\text{degree}(f) \geq 2$ . Glonek (1993) was the first to show this in his study of the case  $f(x) = x_1x_2$  that is relevant, in particular, for hypotheses that are the union of two sets. Moreover, the asymptotic distribution in this case is smaller than  $\chi_1^2$ , making the Wald test maintain (at times quite conservatively) a desired asymptotic level across the entire null hypothesis. We will show that similar phenomena hold also in degree higher than two; see Section 3 that treats monomials  $f(x) = x_1^{\alpha_1}x_2^{\alpha_2}$ . For the tetrad, conservativeness has been remarked upon in work such as Johnson and Bodner (2007). According to our work in Section 5, this is due to the singular nature of the hypothesis rather than effects of too small a sample size. We remark that in singular settings standard  $n$ -out-of- $n$  bootstrap tests may

fail to achieve a desired asymptotic size, requiring the need to consider  $m$ -out-of- $n$  and subsampling procedures; compare the discussion and references in [Drton and Williams \(2011\)](#).

In the remainder of this paper we first clarify the connection between Wald tests and the random variables  $W_{f,\Sigma}$  from (1.1); see Section 2. Bivariate monomials  $f$  of arbitrary degree are the topic of Section 3. Quadratic forms  $f$  are treated in Section 4, which gives a full classification of the bivariate case. The tetrad is studied in Section 5. Our proofs make heavy use of the polar coordinate representation of a pair of independent standard normal random variables and, unfortunately, we have so far not been able to prove the following conjecture, which we discuss further in Section 6.

**CONJECTURE 1.1.** *Let  $\Sigma$  be any positive semidefinite  $k \times k$  matrix with positive diagonal entries. If  $f(x_1, \dots, x_k) = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}$  with nonnegative real exponents  $\alpha_1, \dots, \alpha_k$  that are not all zero, then*

$$W_{f,\Sigma} \sim \frac{1}{(\alpha_1 + \cdots + \alpha_k)^2} \chi_1^2.$$

It is not difficult to show that the conjecture holds when  $\Sigma$  is diagonal.

**PROOF UNDER INDEPENDENCE.** Let  $Z$  be a standard normal random variable, and  $\alpha > 0$ . Then  $\alpha^2/Z^2$  follows the one-sided stable distribution of index  $\frac{1}{2}$  with parameter  $\alpha$ , which has the density

$$(1.3) \quad p_\alpha(x) = \frac{\alpha}{\sqrt{2\pi}} \frac{1}{\sqrt{x^3}} e^{-\frac{1}{2}\alpha^2/x}, \quad x > 0.$$

The law in (1.3) is the distribution of the first passage time of a Brownian motion to the level  $\alpha$  ([Feller, 1966](#)). Hence, it has the convolution rule

$$(1.4) \quad p_\alpha * p_\beta = p_{\alpha+\beta} \quad \alpha, \beta > 0.$$

When  $f(x) = x_1^{\alpha_1} \cdots x_k^{\alpha_k}$  and  $\Sigma = (\sigma_{ij})$  is diagonal with  $\sigma_{11}, \dots, \sigma_{kk} > 0$ , then

$$\frac{1}{W_{f,\Sigma}} = \frac{\alpha_1^2 \sigma_{11}}{X_1^2} + \cdots + \frac{\alpha_k^2 \sigma_{kk}}{X_k^2}.$$

By (1.4), the distribution of  $1/W_{f,\Sigma}$  is  $(\alpha_1 + \cdots + \alpha_k)^2/Z^2$ . Therefore,

$$(1.5) \quad W_{f,\Sigma} \sim \frac{1}{(\alpha_1 + \cdots + \alpha_k)^2} \chi_1^2,$$

as claimed in Conjecture 1.1. □

The preceding argument can be traced back to [Shepp \(1964\)](#); see also [DasGupta and Shepp \(2004\)](#). However, if  $X$  is a dependent random vector, the argument no longer applies. The case  $k = 2$ ,  $\alpha_1 = \alpha_2 = 1$  and  $\Sigma$  arbitrary was proved by [Glonek \(1993\)](#); see Theorem 2.3 below. We prove the general statement for  $k = 2$  as Theorem 3.1.

**2. Wald tests.** To make the connection between Wald tests and the random variables  $W_{f,\Sigma}$  from (1.1) explicit, suppose that  $\theta \in \mathbb{R}^k$  is a parameter of a statistical model and that, based on a sample of size  $n$ , we wish to test the hypothesis

$$(2.1) \quad H_0 : \gamma(\theta) = 0 \quad \text{versus} \quad H_1 : \gamma(\theta) \neq 0$$

for a continuously differentiable function  $\gamma : \mathbb{R}^k \rightarrow \mathbb{R}$ . Suppose further that there is a  $\sqrt{n}$ -consistent estimator  $\hat{\theta}$  of  $\theta$  such that, as  $n \rightarrow \infty$ , we have the convergence in distribution

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}_k(0, \Sigma(\theta)),$$

where the asymptotic covariance matrix  $\Sigma(\theta)$  is a continuous function of the parameter. The Wald statistic for testing (2.1) is the ratio

$$(2.2) \quad T_\gamma = \frac{\gamma(\hat{\theta})^2}{\widehat{\text{var}}[\gamma(\hat{\theta})]} = \frac{n\gamma(\hat{\theta})^2}{(\nabla\gamma(\hat{\theta}))^T \Sigma(\hat{\theta}) \nabla\gamma(\hat{\theta})},$$

where the denominator of the right-most term estimates the asymptotic variance of  $\gamma(\hat{\theta})$ , which by the delta method is given by

$$(\nabla\gamma(\theta))^T \Sigma(\theta) \nabla\gamma(\theta).$$

Consider now a true distribution from  $H_0$ , that is, the true parameter satisfies  $\gamma(\theta) = 0$ . Without loss of generality, we assume that  $\theta = 0$ . If the gradient is nonzero at the true parameter, then the limiting distribution of  $T_\gamma$  is the distribution of the random variable in (1.2) with  $a = \nabla\gamma(0) \neq 0$  and  $\Sigma = \Sigma(0)$ . Hence, the limit is  $\chi_1^2$ . However, if  $\nabla\gamma(0) = 0$  (i.e., the constraint  $\gamma$  is singular at the true parameter), then the asymptotic distribution of  $T_\gamma$  is no longer  $\chi_1^2$  but rather given by (1.1) with the polynomial  $f$  having higher degree; the degree of  $f$  is determined by how many derivatives of  $\gamma$  vanish at the true parameter.

**PROPOSITION 2.1.** *Assume that  $\gamma(0) = 0$  and that there is a homogeneous polynomial  $f$  of degree  $d \geq 2$  such that, as  $x \rightarrow 0$ ,*

$$\gamma(x) = f(x) + o(\|x\|^{d/2}), \quad \text{and} \quad \nabla\gamma(x) = \nabla f(x) + o(\|x\|^{(d-1)/2}).$$

*If  $\sqrt{n}\hat{\theta} \xrightarrow{d} \mathcal{N}(0, \Sigma)$ , then  $T_\gamma \xrightarrow{d} W_{f,\Sigma}$ .*

EXAMPLE 2.2. [Głonek \(1993\)](#) studied testing collapsibility properties of  $2 \times 2 \times 2$  contingency tables. Under an assumption of no three-way interaction, collapsibility with respect to a chosen margin amounts to the vanishing of at least one of two pairwise interactions, which we here simply denote by  $\theta_1$  and  $\theta_2$ . In the  $(\theta_1, \theta_2)$ -plane, the hypothesis is the union of the two coordinate axes, which can be described as the solution set of  $\gamma(\theta_1, \theta_2) = \theta_1\theta_2 = 0$  and tested using the Wald statistic  $T_\gamma$  based on maximum likelihood estimates of  $\theta_1$  and  $\theta_2$ . The hypothesis is singular at the origin as reflected by the vanishing of  $\nabla\gamma$  when  $\theta_1 = \theta_2 = 0$ . Away from the origin,  $T_\gamma$  has the expected asymptotic  $\chi_1^2$  distribution. At the origin, by Proposition 2.1,  $T_\gamma$  converges to  $W_{f,\Sigma}$ , where  $f(x) = x_1x_2$  and  $\Sigma$  is the asymptotic covariance matrix of the two maximum likelihood estimates. The main result of [Głonek \(1993\)](#), stated as a theorem below, gives the distribution of  $W_{f,\Sigma}$  in this case. Głonek's surprising result clarifies that the Wald test for this hypothesis is conservative at (and in finite samples near) the intersection of the two sets making up the null hypothesis.

THEOREM 2.3 ([Głonek, 1993](#)). *If  $f(x) = x_1x_2$  and  $\Sigma$  is any positive semidefinite  $2 \times 2$  matrix with positive diagonal entries, then*

$$W_{f,\Sigma} \sim \frac{1}{4} \chi_1^2.$$

Before turning to concrete problems, we make two simple observations that we will use to bring  $(f, \Sigma)$  in convenient form.

LEMMA 2.4. *Let  $f \in \mathbb{R}[x_1, \dots, x_k]$  be a homogeneous polynomial, and let  $\Sigma$  be a positive semidefinite  $k \times k$  matrix with positive diagonal entries.*

- (i) *If  $c \in \mathbb{R} \setminus \{0\}$  is a nonzero scalar, then  $W_{cf,\Sigma} = W_{f,\Sigma}$ .*
- (ii) *If  $B$  is an invertible  $k \times k$  matrix, then  $W_{f \circ B, B^{-1}\Sigma B^{-T}}$  has the same distribution as  $W_{f,\Sigma}$ .*

PROOF. (i) Obvious, since  $\nabla(cf) = c\nabla f$ . (ii) Let  $X \sim \mathcal{N}(0, \Sigma)$  and define

$$Y = B^{-1}X \sim \mathcal{N}(0, B^{-1}\Sigma B^{-T}).$$

Then  $f(X) = (f \circ B)(Y)$  and  $\nabla(f \circ B)(Y) = B^T \nabla f(X)$ . Substituting into (1.1) gives

$$W_{f,\Sigma} = \frac{(f \circ B)(Y)^2}{(\nabla(f \circ B)(Y))^T B^{-1}\Sigma B^{-T} \nabla(f \circ B)(Y)} = W_{f \circ B, B^{-1}\Sigma B^{-T}}. \quad \square$$

**3. Bivariate Monomials.** In this section, we study the random variable  $W_{f,\Sigma}$  when  $f(x) = x_1^{\alpha_1} x_2^{\alpha_2}$ . If the exponents  $\alpha_1, \alpha_2$  are positive integers, then  $f$  is a bivariate monomial. However, all our arguments go through for a slightly more general case in which  $\alpha_1, \alpha_2$  are positive real numbers. Our main result is that the distribution of  $W_{f,\Sigma}$  does not depend on  $\Sigma$ .

**THEOREM 3.1.** *Let  $f(x) = x_1^{\alpha_1} x_2^{\alpha_2}$  with  $\alpha_1, \alpha_2 > 0$ , and let  $\Sigma$  be any positive semidefinite  $2 \times 2$  matrix with positive diagonal entries. Then*

$$W_{f,\Sigma} \sim \frac{1}{(\alpha_1 + \alpha_2)^2} \chi_1^2.$$

**PROOF.** As shown in Section 1, the claim is true if  $\Sigma = (\sigma_{ij})$  is diagonal. It thus suffices to show that  $W_{f,\Sigma}$  has the same distribution as  $W_f := W_{f,I}$ .

By Lemma 2.4, we can assume without loss of generality that  $\sigma_{11} = \sigma_{22} = 1$  and  $\rho := \sigma_{12} > 0$ . Since

$$\frac{1}{W_{f,\Sigma}} = \frac{\alpha_1^2}{X_1^2} + \frac{2\rho\alpha_1\alpha_2}{X_1X_2} + \frac{\alpha_2^2}{X_2^2},$$

we can also assume  $\alpha_1 = 1$  for simplicity. With  $\sigma = 1/\alpha_2$ , we have

$$\frac{1}{W_{f,\Sigma}} = \frac{1}{X_1^2} + \frac{2\rho}{\sigma X_1 X_2} + \frac{1}{\sigma^2 X_2^2}$$

and need to show that

$$W_{f,\Sigma} \sim \frac{1}{(1 + \sigma^{-1})^2} \chi_1^2.$$

If  $\rho = 1$ , then  $X_1$  and  $X_2$  are almost surely equal and it is clear that  $W_{f,\Sigma}$  has the same distribution as  $W_f$ . Hence, it remains to consider  $0 < \rho < 1$ .

Let  $Z_1$  and  $Z_2$  be independent standard normal random variables. When expressing  $Z_1 = R \cos(\Psi)$  and  $Z_2 = R \sin(\Psi)$  in polar coordinates, it holds that  $R$  and  $\Psi$  are independent, and  $\Psi$  is uniformly distributed over  $[0, 2\pi]$ . Let  $\rho = \sin(\phi)$  with  $0 \leq \phi < \pi/2$ , then the joint distribution of  $X_1$  and  $X_2$  can be represented as

$$X_1 = R \cos(\Psi - \phi/2), \quad X_2 = R \sin(\Psi + \phi/2),$$

which leads to

$$\frac{1}{W_{f,\Sigma}} = \frac{1}{R^2} \cdot \frac{1}{T'}$$

with

$$\frac{1}{T'} = \frac{1}{\cos^2(\Psi - \phi/2)} + \frac{2 \sin(\phi)}{\sigma \cos(\Psi - \phi/2) \sin(\Psi + \phi/2)} + \frac{1}{\sigma^2 \sin^2(\Psi + \phi/2)}.$$

Routine trigonometric calculations show that  $T'$  can be expressed as a function of the doubled angle  $2\Psi$ . More precisely,

$$T' = \frac{\sigma^2}{4} t(2\Psi, \phi),$$

where

$$t(\psi, \phi) = \frac{2 - \cos(2\phi) + 2 \cos(\psi - \phi) - \cos(2\psi) - 2 \cos(\psi + \phi)}{1 - \sigma \cos(2\phi) + (1 + \sigma)[\sigma + \cos(\psi - \phi) - \sigma \cos(\psi + \phi)]}.$$

Since  $2\Psi$  is uniformly distributed on  $[0, 4\pi]$ , the distribution of  $T'$  is independent of  $\phi$  if and only if the same is true for the distribution of  $T = t(\Psi, \phi)$ .

We proceed by calculating the moments of  $T$  and show that they are independent of  $\phi$ . For each  $0 \leq \phi < \pi/2$ , there exists a small interval  $L = [\phi - \epsilon, \phi + \epsilon]$  such that when  $m \geq 1$ , the function

$$\sup_{\phi \in L} [t(\psi, \phi)]^{m-1} \frac{\partial}{\partial \phi} t(\psi, \phi)$$

is integrable over  $0 \leq \psi < 2\pi$ . Therefore, we have

$$(3.1) \quad \frac{\partial}{\partial \phi} \mathbb{E}(T^m) = \int_0^{2\pi} \frac{m}{2\pi} [t(\psi, \phi)]^{m-1} \frac{\partial}{\partial \phi} t(\psi, \phi) d\psi,$$

The expression of  $\frac{\partial}{\partial \phi} t(\psi, \phi)$  is long, so we omit it here.

We introduce the complex numbers  $z = e^{i\psi}$  and  $a = e^{-i\phi}$ , and express the functions  $t(\psi, \phi)$  and  $\frac{\partial}{\partial \phi} t(\psi, \phi)$  in terms of  $z$  and  $a$ :

$$\begin{aligned} t(\psi, \phi) = u(z, a) &= \frac{(a - z)^2(1 + az)^2}{z(a + a\sigma + a^2z - \sigma z)(-1 + a^2\sigma - az - a\sigma z)}, \\ \frac{\partial}{\partial \phi} t(\psi, \phi) = v(z, a) &= \frac{a(a - z)(1 + az)(1 + a^2s + 2az - 2a\sigma z + a^2z^2 + \sigma z^2)}{iz(a + a\sigma + a^2z - \sigma z)^2(-1 + a^2\sigma - az - a\sigma z)^2} \\ &\quad \times (-a - a\sigma - z - a^2z + \sigma z + a^2\sigma z - az^2 - a\sigma z^2). \end{aligned}$$

The integral in (3.1) can be computed as a complex contour integral on the unit circle  $\mathbb{T} = \{z : |z| = 1\}$

$$\int_0^{2\pi} [t(\psi, \phi)]^{m-1} \frac{\partial}{\partial \phi} t(\psi, \phi) d\psi = \oint_{\mathbb{T}} [u(z, a)]^{m-1} v(z, a) \frac{1}{iz} dz.$$

Let

$$q(z, a) = [u(z, a)]^{m-1} v(z, a) \frac{1}{iz}.$$

As a function of  $z$ , it has two poles within the unit disc. These two poles are at  $z_0 = 0$  and  $z_1 = (a^2\sigma - 1)/(a + a\sigma)$  and have the same order  $m + 1$ . By the Residue Theorem, we know

$$(3.2) \quad \frac{1}{2\pi i} \oint_{\mathbb{T}} q(z, a) dz = \text{Res}(q; 0) + \text{Res}(q; z_1),$$

where  $\text{Res}(q; 0)$  and  $\text{Res}(q; z_1)$  are the residues at 0 and  $z_1$  respectively. Let  $\zeta_0 = \{ce^{i\psi}, 0 \leq \psi \leq 2\pi\}$  be a small circle around 0 such that  $z_1$  is outside the circle. Let  $S$  be the Möbius transform

$$S(w) = \frac{z_1 - w}{1 - \bar{z}_1 w}.$$

Then  $S$  is one-to-one from the unit disk onto itself and maps 0 to  $z_1$ , and  $\zeta_0$  to a closed curve  $\zeta_1 = \{S(ce^{i\psi}), 0 \leq \psi \leq 2\pi\}$  around  $z_1$  with winding number one. It holds that

$$\text{Res}(q; z_1) = \frac{1}{2\pi i} \oint_{\zeta_1} q(z, a) dz = \frac{1}{2\pi i} \oint_{\zeta_0} q(S(w), a) S'(w) dw.$$

It also holds that

$$q(S(w), a) S'(w) = -q(w, a),$$

from which we deduce

$$\frac{1}{2\pi i} \oint_{\zeta_0} q(S(w), a) S'(w) dw = -\frac{1}{2\pi i} \oint_{\zeta_0} q(w, a) dw = -\text{Res}(q; 0).$$

Hence, the integral in (3.2) is zero.

We have shown that the integral in (3.1) is zero for every  $m \geq 1$ , which means that the moments of  $T$  do not depend on  $\phi$  for  $0 \leq \phi < \pi/2$ . When  $\phi = 0$ , the random variable  $T$  is bounded, so its moments uniquely determine the distribution. Therefore, the distribution of  $T$  does not depend on  $\phi$ , and the proof is complete.  $\square$

REMARK 3.2. If  $\alpha_1 = \alpha_2$ , then Theorem 3.1 reduces to Theorem 2.3. In this case, our proof above would only need to treat  $\sigma = 1$ . Glonek's proof of Theorem 2.3 finds the distribution function of a random variable related to our  $T$ . If  $\sigma = 1$ , this requires solving a quadratic equation. When  $\sigma \neq 1$ , we were unable to extend this approach as a complicated quartic equation arises in the computation of the distribution function. We thus turned to the presented method of moments.



Let  $X = (X_1, X_2)^T$  and  $Y = (Y_1, Y_2)^T$  be two independent  $\mathcal{N}_2(0, \Sigma)$  random vectors, where  $\Sigma$  has positive diagonal entries. Let  $p_1, p_2$  be nonnegative numbers such that  $p_1 + p_2 = 1$ . The random variable

$$Q = \frac{p_1 X_2 Y_1 + p_2 X_1 Y_2}{\sqrt{(p_1 X_2, p_2 X_1) \Sigma (p_1 X_2, p_2 X_1)^T}}$$

has the standard normal distribution, and is independent of  $X$ . For  $f(x) = x_1^{p_1} x_2^{p_2}$ , let

$$V_{f,\Sigma} = \frac{f(X)}{\sqrt{(\nabla f(X))^T \Sigma \nabla f(X)}}$$

and  $W_{f,\Sigma} = V_{f,\Sigma}^2$ . Then

$$(3.3) \quad p_1 \frac{Y_1}{X_1} + p_2 \frac{Y_2}{X_2} = \frac{Q}{V_{f,\Sigma}}.$$

By taking the conditional expectation given  $V_{f,\Sigma}$ , the characteristic function of (3.3) is seen to be

$$\mathbb{E}[\exp\{itQ/V_{f,\Sigma}\}] = \mathbb{E}[\exp\{-\frac{1}{2}t^2/W_{f,\Sigma}\}].$$

The uniqueness of the moment generating function for positive random variables (Billingsley, 1995, Thm. 22.2) yields that (3.3) has a standard Cauchy distribution (with characteristic function  $e^{-|t|}$ ) if and only if  $W_{f,\Sigma} \sim \chi_1^2$ . Therefore, we have the following equivalent version of Theorem 3.1.

**COROLLARY 3.3.** *Let  $X = (X_1, X_2)^T$  and  $Y = (Y_1, Y_2)^T$  be independent  $\mathcal{N}_2(0, \Sigma)$  random vectors, where  $\Sigma$  has positive diagonal entries. If  $p_1, p_2$  are nonnegative numbers such that  $p_1 + p_2 = 1$ , then the random variable*

$$p_1 \frac{Y_1}{X_1} + p_2 \frac{Y_2}{X_2}$$

*has the standard Cauchy distribution.*

**4. Quadratic Forms.** In this section, we consider the distribution of  $W_{f,\Sigma}$  when  $f$  is a quadratic form, that is,

$$f(x_1, x_2, \dots, x_k) = \sum_{1 \leq i \leq j \leq k} a'_{ij} x_i x_j$$

for real coefficients  $a'_{ij}$ . Equivalently,

$$(4.1) \quad f(x_1, x_2, \dots, x_k) = x^T A x,$$

where  $A = (a_{ij})$  is symmetric, with  $a_{ii} = a'_{ii}$  and  $a_{ij} = a'_{ij}/2$  for  $i < j$ .

4.1. *Canonical form.* Let  $I$  denote the  $k \times k$  identity matrix. We use the shorthand

$$W_f := W_{f,I}$$

when the covariance matrix  $\Sigma$  is the identity.

LEMMA 4.1. *If  $f \in \mathbb{R}[x_1, \dots, x_k]$  is homogeneous of degree  $d$  and  $\Sigma$  is a positive semidefinite  $k \times k$  matrix, then  $W_{f,\Sigma}$  has the same distribution as  $W_g$  where  $g$  is a homogeneous degree  $d$  polynomial in  $\text{rank}(\Sigma)$  many variables.*

PROOF. If  $\Sigma$  has full rank then  $\Sigma = BB^T$  for an invertible matrix  $B$ . Use Lemma 2.4(ii) to transform  $W_{f,\Sigma}$  to  $W_g$  where  $g = f \circ B$  is homogeneous of degree  $d$ . If  $\Sigma$  has rank  $m < k$  then  $\Sigma = BE_mB^T$ , where  $B$  is invertible and  $E_m$  is zero apart from the first  $m$  diagonal entries that are equal to one. Form  $g$  by substituting  $x_{m+1} = \dots = x_k = 0$  into  $f \circ B$ .  $\square$

Further simplifications are possible for a treatment of the random variables  $W_f$ . In the case of quadratic forms, we may restrict attention to canonical forms  $f(x) = \lambda_1 x_1^2 + \dots + \lambda_k x_k^2$ , as shown in the next lemma.

LEMMA 4.2. *Let  $f(x) = x^T A x$  be a quadratic form given by a symmetric  $k \times k$  matrix  $A \neq 0$ . If  $\Sigma$  is a positive definite  $k \times k$  matrix and  $\lambda_1, \dots, \lambda_k$  are the eigenvalues of  $A\Sigma$ , then  $W_{f,\Sigma}$  has the same distribution as*

$$(4.2) \quad \frac{(\lambda_1 Z_1^2 + \dots + \lambda_k Z_k^2)^2}{4(\lambda_1^2 Z_1^2 + \dots + \lambda_k^2 Z_k^2)},$$

where  $Z_1, \dots, Z_k$  are independent standard normal random variables.

PROOF. Write  $\Sigma = BB^T$  for an invertible matrix  $B$ . By Lemma 2.4(ii),  $W_{f,\Sigma}$  has the same distribution as  $W_g$  with  $g(x) = x^T (B^T A B)x$ . Let

$$Q^T (B^T A B) Q = \text{diag}(\lambda_1, \dots, \lambda_k)$$

be the spectral decomposition of  $B^T A B$ , with  $Q$  orthogonal. Then  $\lambda_1, \dots, \lambda_k$  are also the eigenvalues of  $A\Sigma$ . Applying Lemma 2.4(ii) again, we find that  $W_{f,\Sigma}$  has the same distribution as  $W_h$  with

$$h(x) = x^T (Q^T B^T A B Q)x = \lambda_1 x_1^2 + \dots + \lambda_k x_k^2.$$

Since  $\nabla h(x) = 2(\lambda_1 x_1, \dots, \lambda_k x_k)$ , the claim follows.  $\square$

In (4.2), the set of eigenvalues  $\{\lambda_i : 1 \leq i \leq k\}$  can be scaled to  $\{c\lambda_i : 1 \leq i \leq k\}$  for any  $c \neq 0$ , without changing the distribution; recall also Lemma 2.4(i). For instance, we may scale one nonzero eigenvalue to become equal to one. When all (scaled)  $\lambda_i$  are in  $\{-1, 1\}$ , the description of the distribution of  $W_{f,\Sigma}$  can be simplified. We write  $\text{Beta}(\alpha, \beta)$  for the Beta distribution with parameters  $\alpha, \beta > 0$ .

LEMMA 4.3. *Let  $k_1$  and  $k_2$  be two positive integers, and let  $k = k_1 + k_2$ . If  $f(x_1, \dots, x_k) = x_1^2 + \dots + x_{k_1}^2 - x_{k_1+1}^2 - \dots - x_{k_1+k_2}^2$ , then  $W_f$  has the same distribution as*

$$\frac{1}{4}R^2(2B-1)^2,$$

where  $R^2$  and  $B$  are independent,  $R^2 \sim \chi_k^2$ , and  $B \sim \text{Beta}(k_1/2, k_2/2)$ .

PROOF. The distribution of  $W_f$  is that of

$$\frac{1}{4} \frac{(Z_1^2 + \dots + Z_{k_1}^2 - Z_{k_1+1}^2 - \dots - Z_{k_1+k_2}^2)^2}{Z_1^2 + \dots + Z_k^2}$$

with  $Z_1, \dots, Z_k$  independent and standard normal. Let

$$Y_1 := Z_1^2 + \dots + Z_{k_1}^2 \sim \chi_{k_1}^2, \quad Y_2 := Z_{k_1+1}^2 + \dots + Z_k^2 \sim \chi_{k_2}^2.$$

Then  $R^2 := Y_1 + Y_2 \sim \chi_k^2$ . Representing  $Z_1, \dots, Z_k$  in polar coordinates shows that  $R^2$  and  $W_f/R^2$  are independent (Muirhead, 1982, Thm. 1.5.5). Since  $B = Y_1/(Y_1 + Y_2) \sim \text{Beta}(k_1/2, k_2/2)$ , and  $(Y_1 - Y_2)^2/(Y_1 + Y_2)^2 = (2B - 1)^2$ , we deduce that the two random variables  $W_f/R^2$  and  $\frac{1}{4}(2B - 1)^2$  have the same distribution.  $\square$

We note that when  $k = 4$  and  $k_1 = k_2 = 2$ , Lemma 4.3 gives the equality of distributions

$$(4.3) \quad W_f \stackrel{d}{=} \frac{1}{4}R^2U^2.$$

The equality holds because, in this special case,  $U = Y_1/(Y_1 + Y_2)$  is uniformly distributed on  $[0, 1]$ , and  $(2U - 1)^2$  has the same distribution as  $U^2$ . The distribution from (4.3) will appear in Section 5.

For general eigenvalues  $\lambda_i$ , it seems that the distribution from (4.2) cannot be described in as simple terms.

4.2. *Classification of bivariate quadratic forms.* We now turn to the bivariate case ( $k = 2$ ), that is, we are considering a quadratic form in two variables,

$$f(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2.$$

In this case, we are able to give a full classification of the possible distributions of  $W_f$  in terms of linear combinations of a pair of independent  $\chi_1^2$  random variables; see [Johnson, Kotz and Balakrishnan \(1994, Sect. 18.8\)](#) for a discussion of such distributions. Our classification reveals that for  $k = 2$  the distributions for quadratic forms are stochastically bounded below and above by  $\chi_1^2/4$  and  $\chi_2^2/4$ , respectively.

**THEOREM 4.4.** *Let  $\Sigma$  be a positive definite matrix, and let  $f(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2$  be a nonzero quadratic form with matrix*

$$A := \begin{pmatrix} a & b \\ b & c \end{pmatrix} \neq 0.$$

- (a) *If  $b^2 - ac \geq 0$ , then  $W_{f,\Sigma} \sim \chi_1^2/4$ .*
- (b) *If  $b^2 - ac < 0$ , then*

$$W_{f,\Sigma} \stackrel{d}{=} \frac{1}{4} \left( Z_1^2 + \frac{4 \det(A\Sigma)}{\text{tr}(A\Sigma)^2} Z_2^2 \right),$$

*where  $Z_1$  and  $Z_2$  are independent standard normal random variables.*

**PROOF.** (a) When the discriminant  $b^2 - ac \geq 0$  then  $f$  factors into a product of two linear forms. The joint distribution of the two linear forms is bivariate normal. Write  $\Sigma'$  for the covariance matrix of the linear forms then the distribution of  $W_f$  is equal to the distribution of  $W_{g,\Sigma'}$  with  $g(x_1, x_2) = x_1x_2$ . Hence, the distribution is  $\chi_1^2/4$  by [Theorem 2.3/Theorem 3.1](#).

(b) In this case, the discriminant is negative and  $f$  does not factor. By [Lemma 4.2](#), we can assume  $\Sigma = I$  and consider the distribution of  $W_f$  for  $f(x_1, x_2) = \lambda_1 x_1^2 + \lambda_2 x_2^2$ , where  $\lambda_1$  and  $\lambda_2$  are the eigenvalues of  $A\Sigma$ . Since  $\det(A\Sigma) = \lambda_1 \lambda_2$  and  $\text{tr}(A\Sigma) = \lambda_1 + \lambda_2$ , to prove the claim, we must show that in this case

$$(4.4) \quad W_f \stackrel{d}{=} \frac{1}{4} \left( Z_1^2 + \frac{4\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2} Z_2^2 \right) = \frac{1}{4} \left( Z_1^2 + \frac{4c}{(1+c)^2} Z_2^2 \right),$$

where  $c = \lambda_2/\lambda_1 > 0$ .

To show [\(4.4\)](#) we use the polar coordinates again. So represent the two considered independent standard normal random variables as  $X_1 = R \cos(\Psi)$

and  $X_2 = R \sin(\Psi)$ , where  $R^2 \sim \chi_2^2$  and  $\Psi \sim \text{Uniform}[0, 2\pi]$  are independent. Then

$$\begin{aligned} W_f &= \frac{R^2}{4} \cdot \frac{[\cos(\Psi)^2 + c \sin(\Psi)^2]^2}{\cos(\Psi)^2 + c^2 \sin(\Psi)^2} \\ &= \frac{R^2}{4} \cdot \left( 1 - \frac{(1-c)^2}{(1+c)^2} \frac{\cos(\Psi)^2 \sin(\Psi)^2}{\cos(\Psi)^2 + c^2 \sin(\Psi)^2} \right). \end{aligned}$$

Using Lemma 4.5, we have

$$\begin{aligned} (4.5) \quad W_f &\stackrel{d}{=} \frac{R^2}{4} \left( 1 - \frac{(1-c)^2}{(1+c)^2} \cos(\Psi)^2 \right) \\ &= \frac{R^2}{4} \left( \frac{4c}{(1+c)^2} \cos(\Psi)^2 + \sin(\Psi)^2 \right). \end{aligned}$$

This is the claim from (4.4) because  $R \cos(\Psi)$  and  $R \sin(\Psi)$  are independent and standard normal.  $\square$

LEMMA 4.5. *If  $c \geq 0$  and  $\Psi$  has a uniform distribution over  $[0, 2\pi]$ , then*

$$S_c(\Psi) := \frac{(1+c)^2 \cos(\Psi)^2 \sin(\Psi)^2}{\cos(\Psi)^2 + c^2 \sin(\Psi)^2} \stackrel{d}{=} \cos(\Psi)^2.$$

PROOF. Let  $R^2 \sim \chi_2^2$  be independent of  $\Psi$ . Then  $R \sin(\Psi)$  and  $R \cos(\Psi)$  are independent and standard normal. Therefore,

$$\frac{1}{R^2 S_c(\Psi)} = \frac{1}{(1+c)^2} \cdot \frac{1}{[R \sin(\Psi)]^2} + \frac{c^2}{(1+c)^2} \frac{1}{[R \cos(\Psi)]^2}$$

is the sum of two independent random variables that follow the one-sided stable distribution of index  $\frac{1}{2}$ . Since  $c > 0$ , the first summand has the stable distribution with parameter  $1/(1+c)$  and the second summand has parameter  $c/(1+c)$ . Hence, by (1.4), their sum follows a stable law with parameter 1. Expressing this in terms of the reciprocals,

$$R^2 S_c(\Psi) \stackrel{d}{=} R^2 \cos(\Psi)^2 \sim \chi_1^2.$$

It follows that  $S_c(\Psi)$  has the same distribution as  $\cos(\Psi)^2$ . For instance, we may argue that  $S_c(\Psi)$  and  $\cos(\Psi)^2$  have identical moments, which implies equality of the distributions as both are compactly supported.  $\square$

We remark that the claim of Lemma 4.5 is false for  $c < 0$ . Indeed, the distribution of  $S_c(\Psi)$  varies with  $c$  when  $c < 0$ .

4.3. *Stochastic bounds.* To understand possible conservativeness of Wald tests it is interesting to look for stochastic bounds on  $W_{f,\Sigma}$  that hold for all  $f$  and  $\Sigma$ . We denote the stochastic ordering of two random variables as  $U \leq_{st} V$  when  $P(U > t) \leq P(V > t)$  for all  $t \in \mathbb{R}$ .

PROPOSITION 4.6. *If  $f \in \mathbb{R}[x_1, \dots, x_k]$  is a quadratic form and  $\Sigma$  any nonzero positive semidefinite  $k \times k$  matrix, then  $W_{f,\Sigma} \leq_{st} \frac{1}{4}\chi_k^2$ . Equality is achieved when  $f(x) = x_1^2 + \dots + x_k^2$  and  $\Sigma$  is the identity matrix.*

PROOF. The second claim is obvious. For the first claim, without loss of generality, we can restrict our attention to the distributions from (4.2). The Cauchy-Schwarz inequality gives

$$\begin{aligned} \frac{(\lambda_1 Z_1^2 + \dots + \lambda_k Z_k^2)^2}{4(\lambda_1^2 Z_1^2 + \dots + \lambda_k^2 Z_k^2)} &\leq \frac{(Z_1^2 + \dots + Z_k^2)(\lambda_1^2 Z_1^2 + \dots + \lambda_k^2 Z_k^2)}{4(\lambda_1^2 Z_1^2 + \dots + \lambda_k^2 Z_k^2)} \\ &= \frac{1}{4}(Z_1^2 + \dots + Z_k^2), \end{aligned}$$

which is the desired chi-square bound.  $\square$

The considered Wald test rejects the hypothesis that  $\gamma(\theta) = 0$  when the statistic  $T_\gamma$  from (2.2) exceeds  $c_\alpha$ , where  $c_\alpha$  is the  $(1 - \alpha)$  quantile of the  $\chi_1^2$  distribution. Let  $k_\alpha$  be the largest degrees of freedom  $k$  such that a  $\frac{1}{4}\chi_k^2$  random variable exceeds  $c_\alpha$  with probability at most  $\alpha$ . According to Proposition 4.6, if the true parameter is a singularity at which  $\gamma$  can be approximated by a quadratic form in at most  $k_\alpha$  variables, then the Wald test is guaranteed to be asymptotically conservative. Some values are

$$k_{0.05} = 7, \quad k_{0.025} = 11, \quad k_{0.01} = 16, \quad k_{0.005} = 20, \quad k_{0.001} = 29.$$

Turning to a lower bound, we can offer the following simple observation.

PROPOSITION 4.7. *Suppose the quadratic form  $f$  is given by a symmetric  $k \times k$  matrix  $A \neq 0$ , and suppose that  $\Sigma$  is a positive definite  $k \times k$  matrix such that all eigenvalues of  $A\Sigma$  are nonnegative. Then  $W_{f,\Sigma} \geq_{st} \frac{1}{4}\chi_1^2$ .*

PROOF. Let  $\lambda_1, \dots, \lambda_k \geq 0$  be the eigenvalues of  $A\Sigma$ . By scaling, we can assume without loss of generality that  $\lambda_1 = 1$  and  $0 \leq \lambda_i \leq 1$  for  $2 \leq i \leq k$ . Then

$$\frac{(\lambda_1 Z_1^2 + \dots + \lambda_k Z_k^2)^2}{4(\lambda_1^2 Z_1^2 + \dots + \lambda_k^2 Z_k^2)} \geq \frac{Z_1^2(\lambda_1^2 Z_1^2 + \dots + \lambda_k^2 Z_k^2)}{4(\lambda_1^2 Z_1^2 + \dots + \lambda_k^2 Z_k^2)} = \frac{1}{4}Z_1^2,$$

and the claim follows from Lemma 4.2.  $\square$

Proposition 4.7, Theorem 4.4 and simulation experiments lead us to conjecture that  $\frac{1}{4}\chi_1^2$  is still a stochastic lower bound when there are both positive and negative eigenvalues  $\lambda_i$ .

CONJECTURE 4.8. *For any quadratic form  $f \neq 0$  and any positive semidefinite matrix  $\Sigma \neq 0$ , the distribution of  $W_{f,\Sigma}$  stochastically dominates  $\frac{1}{4}\chi_1^2$ .*

While we do not know how to prove this conjecture in general, we are able to treat the special case where the eigenvalues  $\lambda_i$  are either 1 or -1.

THEOREM 4.9. *Let  $k_1, k_2 > 0$ , and  $k = k_1 + k_2$ . If  $f(x_1, \dots, x_k) = x_1^2 + \dots + x_{k_1}^2 - x_{k_1+1}^2 - \dots - x_{k_1+k_2}^2$ , then  $W_f \geq_{st} \frac{1}{4}\chi_1^2$ .*

PROOF. Without loss of generality we assume  $k_1 \leq k_2$ . If  $k_1 = 0$  or  $k_1 = k_2 = 1$ , the claim follows Proposition 4.7 and Theorem 4.4, respectively. We now consider the case  $k_1 \geq 1$  and  $k_2 \geq 2$ . By Lemma 4.3, we know

$$(4.6) \quad W_f \stackrel{d}{=} \frac{1}{4}R^2(2B-1)^2,$$

where  $R^2$  and  $B$  are independent,  $R^2 \sim \chi_k^2$ , and  $B \sim \text{Beta}(k_1/2, k_2/2)$ . On the other hand, if  $B' \sim \text{Beta}(1/2, (k-1)/2)$  and is independent of  $R^2$ , then

$$(4.7) \quad R^2 B' \sim \chi_1^2.$$

Let  $g(x)$  and  $h(x)$  be the density functions of  $(2B-1)^2$  and  $B'$ , respectively. The comparison of (4.6) and (4.7) shows that it suffices to prove that  $(2B-1)^2$  is stochastically larger than  $B'$ . We will show a stronger result, namely, that the likelihood ratio  $g(x)/h(x)$  is an increasing function over  $[0, 1]$ .

To simplify the argument, we rescale the density functions to

$$\begin{aligned} g(x)\sqrt{x} &\propto (1+\sqrt{x})^{k_1/2-1}(1-\sqrt{x})^{k_2/2-1} \\ &\quad + (1-\sqrt{x})^{k_1/2-1}(1+\sqrt{x})^{k_2/2-1} \end{aligned}$$

and

$$\begin{aligned} h(x)\sqrt{x} &\propto (1-x)^{(k-3)/2} \\ &= (1-\sqrt{x})^{(k_1+k_2-3)/2}(1+\sqrt{x})^{(k_1+k_2-3)/2}. \end{aligned}$$

For our purpose, it is equivalent to show the monotonicity of  $g(x^2)/h(x^2)$ , which is proportional to

$$\begin{aligned} \ell(x) &:= (1+x)^{(-k_2+1)/2}(1-x)^{(-k_1+1)/2} \\ &\quad + (1-x)^{(-k_2+1)/2}(1+x)^{(-k_1+1)/2}. \end{aligned}$$

When  $k_1 = 1$ , the derivative of  $\ell(x)$  satisfies

$$2\ell'(x) = (k_2 - 1)(1 - x)^{(-k_2-1)/2} - (k_2 - 1)(1 + x)^{(-k_2-1)/2} > 0$$

when  $0 < x < 1$ , and thus the likelihood ratio is an increasing function. When  $k_1 \geq 2$ , we have

$$\begin{aligned} 2\ell'(x) &= (1+x)^{(k_2+1)/2}(1-x)^{(k_2+1)/2} \\ &= (1+x) \left[ (k_2-1)(1+x)^{(k_2-k_1)/2} + (k_1-1)(1-x)^{(k_2-k_1)/2} \right] \\ &\quad - (1-x) \left[ (k_1-1)(1+x)^{(k_2-k_1)/2} + (k_2-1)(1-x)^{(k_2-k_1)/2} \right] \\ &> (k_2 - k_1) \left[ (1+x)^{(k_2-k_1)/2} - (1-x)^{(k_2-k_1)/2} \right] \geq 0 \end{aligned}$$

for all  $0 < x < 1$ . Therefore,  $\ell(x)$  is an increasing function.  $\square$

**5. Tetrads.** We now turn to the problem that sparked our interest in Wald tests of singular hypothesis, namely, the problem of testing tetrad constraints on the covariance matrix  $\Theta = (\theta_{ij})$  of a random vector  $Y$  in  $\mathbb{R}^p$  with  $p \geq 4$ . A tetrad is a  $2 \times 2$  subdeterminant that only involves off-diagonal entries and, without loss of generality, we consider the tetrad

$$(5.1) \quad \gamma(\Theta) = \theta_{13}\theta_{24} - \theta_{14}\theta_{23} = \det \begin{pmatrix} \theta_{13} & \theta_{14} \\ \theta_{23} & \theta_{24} \end{pmatrix}.$$

**EXAMPLE 5.1.** Consider a factor analysis model in which the coordinates of  $Y$  are linear functions of a latent variable  $X$  and noise terms. More precisely,  $Y_i = \beta_{0i} + \beta_i X + \epsilon_i$  where  $X \sim \mathcal{N}(0, 1)$  is independent of  $\epsilon_1, \dots, \epsilon_p$ , which in turn are independent normal random variables. Then the covariance between  $Y_i$  and  $Y_j$  is  $\theta_{ij} = \beta_i \beta_j$  and the tetrad from (5.1) vanishes.

Suppose now that we observe a sample of independent and identically distributed random vectors  $Y^{(1)}, \dots, Y^{(n)}$  with covariance matrix  $\Theta$ . Let  $\bar{Y}_n$  be the sample mean vector, and let

$$\hat{\Theta} = \frac{1}{n} \sum_{i=1}^n (Y^{(i)} - \bar{Y}_n)(Y^{(i)} - \bar{Y}_n)^T$$

be the empirical covariance matrix. Assuming that the data-generating distribution has finite fourth moments, it holds that

$$\sqrt{n}(\hat{\Theta} - \Theta) \xrightarrow{d} \mathcal{N}_k(0, V(\Theta))$$



with  $k = p^2$ . The rows and columns of the asymptotic covariance matrix  $V(\Theta)$  are indexed by the pairs  $ij := (i, j)$ ,  $1 \leq i, j \leq p$ . Since the tetrad from (5.1) only involves the covariances indexed by the pairs in  $C = \{13, 14, 23, 24\}$ , only the principal submatrix

$$\Sigma(\Theta) := V(\Theta)_{C \times C}$$

is of relevance for the large-sample distribution of the sample tetrad  $\gamma(\hat{\Theta})$ .

The gradient of the tetrad is

$$\nabla \gamma(\Theta) = (\theta_{24}, -\theta_{23}, -\theta_{14}, \theta_{13}).$$

Hence, if at least one of the four covariances in the tetrad is nonzero the Wald statistic  $T_\gamma$  converges to a  $\chi_1^2$  distribution. If, on the other hand,  $\theta_{13} = \theta_{14} = \theta_{23} = \theta_{24} = 0$ , then the large-sample limit of  $T_\gamma$  has the distribution of  $W_{f, \Sigma(\Theta)}$  where

$$f(x) = x_1 x_4 - x_2 x_3$$

is a quadratic form in  $k = 4$  variables; recall Proposition 2.1. This form can be written as  $x^T A x$  with a matrix that is a Kronecker product, namely

$$(5.2) \quad A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

If  $Y$  is multivariate normal, then the asymptotic covariance matrix has the entries

$$V(\Theta)_{ij,kl} = \theta_{ik}\theta_{jl} + \theta_{il}\theta_{jk}.$$

In the singular case with  $\theta_{13} = \theta_{14} = \theta_{23} = \theta_{24} = 0$ , we have thus

$$\Sigma(\Theta) = \begin{pmatrix} \theta_{11}\theta_{33} & \theta_{11}\theta_{34} & \theta_{12}\theta_{33} & \theta_{12}\theta_{34} \\ \theta_{11}\theta_{34} & \theta_{11}\theta_{44} & \theta_{12}\theta_{34} & \theta_{12}\theta_{44} \\ \theta_{12}\theta_{33} & \theta_{12}\theta_{34} & \theta_{22}\theta_{33} & \theta_{22}\theta_{34} \\ \theta_{12}\theta_{34} & \theta_{12}\theta_{44} & \theta_{22}\theta_{34} & \theta_{22}\theta_{44} \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{pmatrix} \otimes \begin{pmatrix} \theta_{33} & \theta_{34} \\ \theta_{34} & \theta_{44} \end{pmatrix},$$

which again is a Kronecker product. We remark that  $\Sigma(\Theta)$  would also be a Kronecker product if we had started with an elliptical distribution instead of the normal, compare Iwashita and Siotani (1994, eqn. (2.1)), or if  $(Y_1, Y_2)$  and  $(Y_3, Y_4)$  were independent in the data-generating distribution.

As we show next, in the singular case, the Kronecker structure of the two matrices  $A$  and  $\Sigma(\Theta)$  gives a limiting distribution of the Wald statistic for the tetrad that does not depend on the block-diagonal covariance matrix  $\Theta$ .

THEOREM 5.2. *Let  $\Sigma = \Sigma^{(1)} \otimes \Sigma^{(2)}$  be the Kronecker product of two positive definite  $2 \times 2$  matrices  $\Sigma^{(1)}, \Sigma^{(2)}$ . Let  $f(x) = x_1x_4 - x_2x_3$ . Then*

$$W_{f,\Sigma} \stackrel{d}{=} \frac{1}{4}R^2U^2,$$

where  $R^2 \sim \chi_4^2$  and  $U \sim \text{Uniform}[0, 1]$  are independent.

PROOF. Since  $f$  is a quadratic form we may consider the canonical form from Lemma 4.2, which depends on the (real) eigenvalues of  $A\Sigma$ . The claim follows from Lemma 4.3 and the comments in the paragraph following its proof provided the four eigenvalues of  $A\Sigma$  all have the same absolute value, two of them are positive and two are negative.

Let  $\Sigma^{(i)} = (\sigma_{kl}^{(i)})$ . Then, by (5.2),

$$A\Sigma = \begin{pmatrix} -\sigma_{12}^{(1)} & \sigma_{11}^{(1)} \\ -\sigma_{22}^{(1)} & \sigma_{12}^{(1)} \end{pmatrix} \otimes \begin{pmatrix} -\sigma_{12}^{(2)} & \sigma_{11}^{(2)} \\ -\sigma_{22}^{(2)} & \sigma_{12}^{(2)} \end{pmatrix}.$$

For  $i = 1, 2$ , since  $\Sigma^{(i)}$  is positive definite, the matrix

$$\begin{pmatrix} -\sigma_{12}^{(i)} & \sigma_{11}^{(i)} \\ -\sigma_{22}^{(i)} & \sigma_{12}^{(i)} \end{pmatrix}$$

has the imaginary eigenvalues

$$\pm \lambda^{(i)} = \pm \sqrt{(\sigma_{12}^{(i)})^2 - \sigma_{11}^{(i)}\sigma_{22}^{(i)}}.$$

It follows that  $A\Sigma$  has the real eigenvalues

$$\lambda^{(1)}\lambda^{(2)} \quad \text{and} \quad -\lambda^{(1)}\lambda^{(2)},$$

each with multiplicity two. Hence, Lemma 4.3 applies with  $k_1 = k_2 = 2$ .  $\square$

The distribution function of  $\frac{1}{4}R^2U^2$  is

$$F_{\text{sing}}(t) = 1 - e^{-2t} + \sqrt{2\pi t} \left( 1 - \Phi(2\sqrt{t}) \right), \quad t \geq 0,$$

where  $\Phi(t)$  is the distribution function of  $\mathcal{N}(0, 1)$ . The density  $f_{\text{sing}}(t)$  of  $\frac{1}{4}R^2U^2$  is strictly decreasing on  $(0, \infty)$  and  $f_{\text{sing}}(t) \rightarrow \infty$  as  $t \rightarrow 0$ . In light of Theorem 4.4, it is interesting to note that the distribution of  $\frac{1}{4}R^2U^2$  is not the distribution of a linear combination of four independent  $\chi_1^2$  random

variables, because the  $\chi_d^2$  distribution has a finite density at zero when  $d \geq 2$ . However, the distribution satisfies

$$\frac{1}{4}\chi_1^2 \leq_{st} \frac{1}{4}R^2U^2 \leq_{st} \frac{1}{4}\chi_2^2.$$

The first inequality holds according to Theorem 4.9. The second inequality holds because  $R^2U \sim \chi_2^2$ . According to the next result, the distribution is also no larger than a  $\chi_1^2$  distribution, which means that the Wald test of a tetrad constraint is asymptotically conservative at the tetrad's singularities (which are given by block-diagonal covariance matrices).

**PROPOSITION 5.3.** *Suppose  $R^2 \sim \chi_4^2$  and  $U \sim \text{Uniform}[0, 1]$  are independent. Then*

$$\frac{1}{4}R^2U^2 \leq_{st} \chi_1^2.$$

**PROOF.** Let  $Z_1, \dots, Z_4$  be independent standard normal random variables. Then the sum of squares

$$Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 \stackrel{d}{=} R^2 \sim \chi_4^2$$

and the ratio

$$\frac{Z_1^2}{Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2} \sim \text{Beta}\left(\frac{1}{2}, \frac{3}{2}\right)$$

are independent. Hence, the claim holds if and only if

$$\frac{1}{2}U \leq_{st} \sqrt{B},$$

where  $U \sim \text{Uniform}[0, 1]$  and  $B \sim \text{Beta}\left(\frac{1}{2}, \frac{3}{2}\right)$ . The distribution of  $U/2$  is supported on the interval  $[0, 1/2]$  on which it has distribution function

$$F_{U/2}(t) = 2t.$$

For  $t \in (0, 1)$ , the distribution function of  $\sqrt{B}$  has first and second derivative

$$F'_{\sqrt{B}}(t) = \frac{4\sqrt{1-t^2}}{\pi} \quad \text{and} \quad F''_{\sqrt{B}}(t) = -\frac{4t}{\pi\sqrt{1-t^2}}.$$

Hence,  $F_{\sqrt{B}}$  is strictly concave on  $(0, 1)$  and has a tangent with slope  $4/\pi < 2$  at  $t = 0$ . Consequently,

$$F_{U/2}(t) \geq F_{\sqrt{B}}(t), \quad t \in \mathbb{R},$$

giving the claimed ordering of  $\frac{1}{4}R^2U^2$  and the  $\chi_1^2$  distribution.  $\square$

**6. Conjectures.** In Section 3, we mentioned that Theorem 3.1 and Corollary 3.3 are equivalent. Similarly, Conjecture 1.1 is equivalent to the following one.

CONJECTURE 6.1. *Let  $X = (X_1, X_2, \dots, X_k)^T$  and  $Y = (Y_1, Y_2, \dots, Y_k)^T$  be independent and have the same distribution  $\mathcal{N}_k(0, \Sigma)$ , where  $\Sigma$  has positive diagonal entries. If  $p_1, p_2, \dots, p_k$  are nonnegative numbers such that  $p_1 + p_2 + \dots + p_k = 1$ , then*

$$\frac{p_1 Y_1}{X_1} + \frac{p_2 Y_2}{X_2} + \dots + \frac{p_k Y_k}{X_k}$$

*has the standard Cauchy distribution.*

For a proof of this conjecture it is natural to try an induction type argument, which might involve the ratio of normal random variables with nonzero means (Marsaglia, 1965). However, we were unable to make this work.

By taking the reciprocal of  $W_{f,\Sigma}$ , we can translate Conjecture 1.1 into another equivalent form.

CONJECTURE 6.2. *Let  $X = (X_1, X_2, \dots, X_k)^T \sim \mathcal{N}_k(0, \Sigma)$  be such that non of its entries is a point mass. If  $p_1, p_2, \dots, p_n$  are nonnegative numbers such that  $p_1 + p_2 + \dots + p_n = 1$ , then*

$$(6.1) \quad \left( \frac{p_1}{X_1}, \frac{p_2}{X_2}, \dots, \frac{p_n}{X_n} \right) \Sigma \left( \frac{p_1}{X_1}, \frac{p_2}{X_2}, \dots, \frac{p_n}{X_n} \right)^T \sim \frac{1}{\chi_1^2}.$$

Simulation provides strong evidence for the validity of these conjectures. We have tried many randomly generated scenarios with  $2 \leq k \leq 5$ , simulating large numbers of values for the rational functions in question. In all cases empirical distribution functions were indistinguishable from the conjectured  $\chi_1^2$  or Cauchy distribution functions.

On the other hand, the positivity requirement for  $p_1, p_2, \dots, p_k$  is crucial for the validity of the conjectures. For instance, let  $Q$  be the reciprocal of the quantity on the left hand side of (6.1), and consider the special case where  $k = 2$ ,  $\text{var}(X_1) = \text{var}(X_2) = 1$ ,  $\text{cor}(X_1, X_2) = \rho$ , and  $p_1 = -p_2 = 1/2$ . Assuming that  $|\rho| < 1$ , change coordinates to

$$Z_1 = (X_1 + X_2)/\sqrt{2(1 + \rho)}, \quad Z_2 = (X_1 - X_2)/\sqrt{2(1 - \rho)},$$

and then to polar coordinates  $Z_1 = R \cos \Psi$  and  $Z_2 = R \sin \Psi$ . We obtain that

$$Q = 4 \left( \frac{1}{X_1^2} - \frac{2\rho}{X_1 X_2} + \frac{1}{X_2^2} \right)^{-1} = R^2 \frac{[\rho + \cos(2\Psi)]^2}{1 - \rho^2}.$$

The distribution of  $Q$  now depends on  $\rho$ . For instance,

$$\mathbb{E}[Q] = \frac{1 + 2\rho^2}{1 - \rho^2}.$$

**7. Conclusion.** In regular settings, the Wald statistic for testing a constraint on the parameters of a statistical model converges to a  $\chi_1^2$  distribution as the sample size increases. When the true parameter is a singularity of the constraint, the limiting distribution is instead determined by a rational function of jointly normal random variables (recall Section 2). The distributions of these rational functions are in surprising ways related to chi-square distributions as we showed in our main results in Sections 3–5.

Our work led to several, in our opinion, intriguing conjectures about the limiting distributions of Wald statistics. Although the conjectures can be stated in elementary terms, we are not aware of any other work that suggests these properties for the multivariate normal distribution.

For quadratic forms, the usual canonical form leads to a particular class of distributions parametrized by a collection of eigenvalues (recall Lemma 4.2). It would be interesting to study Schur convexity properties of this class of distributions, which would provide further insights into asymptotic conservativeness of Wald tests of singular hypotheses.

Finally, this paper has focused on testing a single constraint. It would be interesting to develop a general theory for Wald tests of hypotheses that are defined in terms of several constraints. In this setting the choice of the constraints representing a null hypothesis will play an important role in the distribution theory, as exemplified by Gaffke, Steyer and von Davier (1999) and Gaffke, Heiligers and Offinger (2002).

**Acknowledgments.** This work was supported by NSF under Grant No. DMS-0746265. Mathias Drton was also supported by an Alfred P. Sloan Fellowship.

## REFERENCES

- AZAÏS, J.-M., GASSIAT, É. and MERCADIER, C. (2006). Asymptotic distribution and local power of the log-likelihood ratio test for mixtures: bounded and unbounded cases. *Bernoulli* **12** 775–799. [MR2265342 \(2008c:62017\)](#)
- BILLINGSLEY, P. (1995). *Probability and measure*, third ed. *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication. [MR1324786 \(95k:60001\)](#)
- BOLLEN, K. A., LENNOX, R. D. and DAHLY, D. L. (2009). Practical application of the vanishing tetrad test for causal indicator measurement models: an example from health-related quality of life. *Stat. Med.* **28** 1524–1536. [MR2649710](#)

- BOLLEN, K. A. and TING, K.-F. (2000). A tetrad test for causal indicators. *Psychological Methods* **5** 3–22.
- CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statistics* **25** 573–578. [MR0065087 \(16,381k\)](#)
- DASGUPTA, A. and SHEPP, L. (2004). Chebyshev polynomials and  $G$ -distributed functions of  $F$ -distributed variables. In *A festschrift for Herman Rubin. IMS Lecture Notes Monogr. Ser.* **45** 153–163. Inst. Math. Statist., Beachwood, OH. [MR2126894 \(2006a:60017\)](#)
- DRTON, M. (2009). Likelihood ratio tests and singularities. *Ann. Statist.* **37** 979–1012. [MR2502658 \(2010d:62143\)](#)
- DRTON, M., MASSAM, H. and OLKIN, I. (2008). Moments of minors of Wishart matrices. *Ann. Statist.* **36** 2261–2283. [MR2458187 \(2010a:60035\)](#)
- DRTON, M., STURMFELS, B. and SULLIVANT, S. (2007). Algebraic factor analysis: tetrads, pentads and beyond. *Probab. Theory Related Fields* **138** 463–493. [MR2299716 \(2008f:62086\)](#)
- DRTON, M., STURMFELS, B. and SULLIVANT, S. (2009). *Lectures on algebraic statistics. Oberwolfach Seminars* **39**. Birkhäuser Verlag, Basel. [MR2723140 \(2012d:62004\)](#)
- DRTON, M. and WILLIAMS, B. (2011). Quantifying the failure of bootstrap likelihood ratio tests. *Biometrika* **98** 919–934. [MR2860333 \(2012k:62065\)](#)
- FELLER, W. (1966). *An introduction to probability theory and its applications. Vol. II*. John Wiley & Sons Inc., New York. [MR0210154 \(35 ##1048\)](#)
- GAFFKE, N., HEILIGERS, B. and OFFINGER, R. (2002). On the asymptotic null-distribution of the Wald statistic at singular parameter points. *Statist. Decisions* **20** 379–398. [MR1960710 \(2004a:62055\)](#)
- GAFFKE, N., STEYER, R. and VON DAVIER, A. A. (1999). On the asymptotic null-distribution of the Wald statistic at singular parameter points. *Statist. Decisions* **17** 339–358. [MR1745064 \(2000j:62027\)](#)
- GLONEK, G. F. V. (1993). On the behaviour of Wald statistics for the disjunction of two regular hypotheses. *J. Roy. Statist. Soc. Ser. B* **55** 749–755. [MR1223941 \(94g:62122\)](#)
- HARMAN, H. H. (1976). *Modern factor analysis*, Third ed. University of Chicago Press, Chicago, Ill. [MR0400546 \(53 ##4377\)](#)
- HIPP, J. R. and BOLLEN, K. A. (2003). Model fit in structural equation models with censored, ordinal, and dichotomous variables: testing vanishing tetrads. *Sociological Methodology* **33** 267–305.
- IWASHITA, T. and SIOTANI, M. (1994). Asymptotic distributions of functions of a sample covariance matrix under the elliptical distribution. *Canad. J. Statist.* **22** 273–283. [MR1295693 \(96b:62086\)](#)
- JOHNSON, T. R. and BODNER, T. E. (2007). A note on the use of bootstrap tetrad tests for covariance structures. *Struct. Equ. Model.* **14** 113–124. [MR2339072](#)
- JOHNSON, N. L., KOTZ, S. and BALAKRISHNAN, N. (1994). *Continuous univariate distributions. Vol. 1*, second ed. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication. [MR1299979 \(96j:62028\)](#)
- KATO, N. and KURIKI, S. (2013). Likelihood ratio tests for positivity in polynomial regressions. *J. Multivariate Anal.* **115** 334–346. [MR3004562](#)
- MARSAGLIA, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *J. Amer. Statist. Assoc.* **60** 193–204. [MR0178490 \(31 ##2747\)](#)
- MUIRHEAD, R. J. (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York. *Wiley Series in Probability and Mathematical Statistics*. [MR652932 \(84c:62073\)](#)

- RITZ, C. and SKOVGAARD, I. M. (2005). Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika* **92** 507–517. [MR2202642 \(2008f:62032\)](#)
- SHEPP, L. (1964). Normal functions of normal random variables. *SIAM Rev.* **6** 459–460.
- SILVA, R., SCHEINES, R., GLYMOUR, C. and SPIRITES, P. (2006). Learning the structure of linear latent variable models. *J. Mach. Learn. Res.* **7** 191–246. [MR2274367](#)
- SPEARMAN, C. (1904). ‘General intelligence,’ objectively determined and measured. *The American Journal of Psychology* **15** 201–292.
- SPIRITES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, prediction, and search*, second ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper and Thomas Richardson, A Bradford Book. [MR1815675 \(2001j:62009\)](#)
- SULLIVANT, S., TALASKA, K. and DRAISMA, J. (2010). Trek separation for Gaussian graphical models. *Ann. Statist.* **38** 1665–1685. [MR2662356 \(2011f:62076\)](#)
- ZWIERNIK, P. and SMITH, J. Q. (2012). Tree cumulants and the geometry of binary tree models. *Bernoulli* **18** 290–321. [MR2888708](#)

DEPARTMENT OF STATISTICS  
UNIVERSITY OF WASHINGTON  
SEATTLE, WA, U.S.A.  
E-MAIL: [md5@uw.edu](mailto:md5@uw.edu)

DEPARTMENT OF STATISTICS & BIOSTATISTICS  
RUTGERS UNIVERSITY  
PISCATAWAY, NJ, U.S.A.  
E-MAIL: [hxiao@stat.rutgers.edu](mailto:hxiao@stat.rutgers.edu)